Chand, M. K., Nirwan, N., Diffin, F. M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M. D., & Saikrishnan, K. (2015). Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes. *Nature Chemical Biology*, *11*, 870-877. https://doi.org/10.1038/nchembio.1926

Supplementary Information

Translocation-coupled DNA cleavage by Type ISP restriction-modification enzymes

Mahesh Kumar Chand[1], Neha Nirwan[1], Fiona M. Diffin[2], Kara van Aelst[2], Manasi Kulkarni[1], Christian Pernstich[2], Mark D. Szczelkun[2*], Kayarat Saikrishnan[1*]

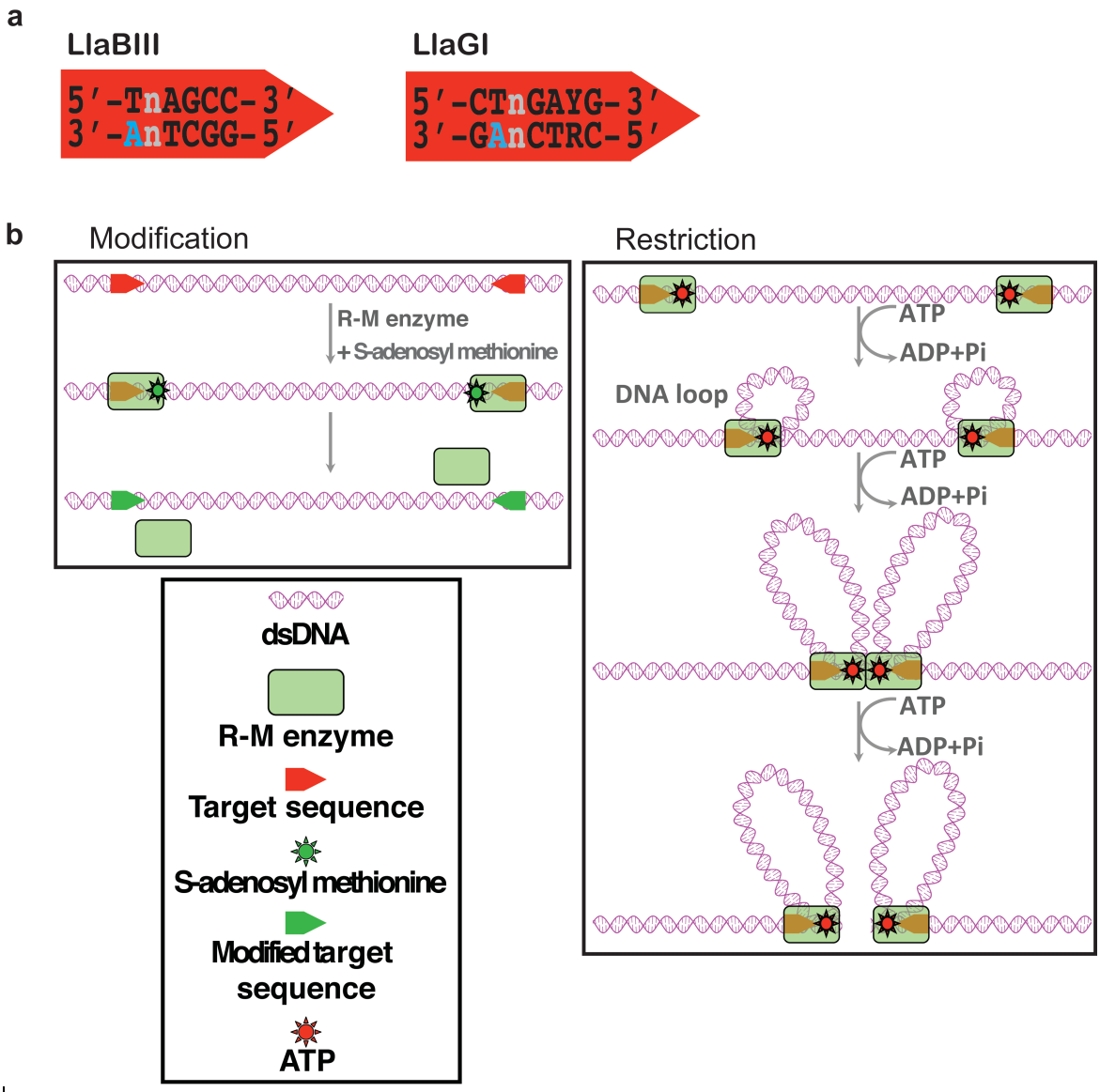
[1]Division of Biology, Indian Institute of Science Education and Research, Pune, 411008, India

[2]DNA-Protein Interactions Unit, School of Biochemistry, Biomedical Sciences Building, University of Bristol, Bristol BS8 1TD, UK

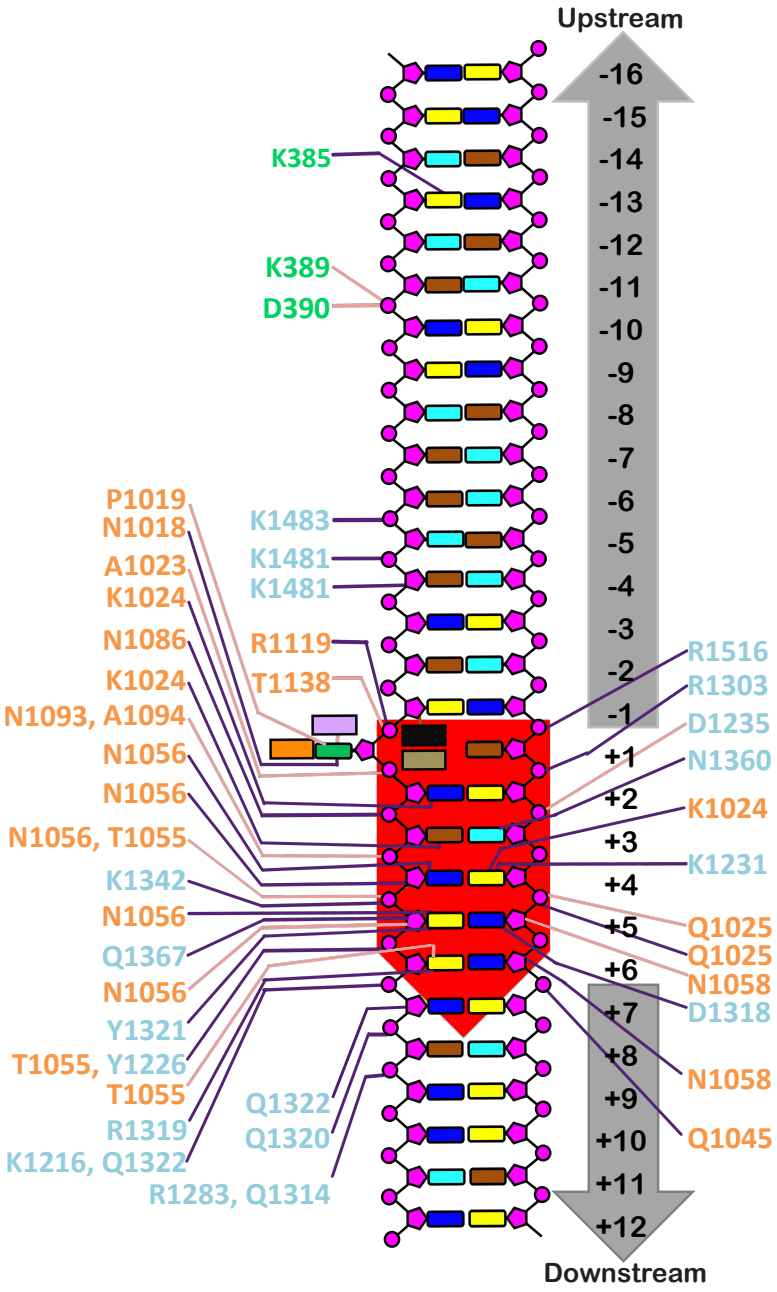
[*]Corresponding authors

Phone: +91 2025908047
Fax: +91 2025908186
email: saikrishnan@iiserpune.ac.in

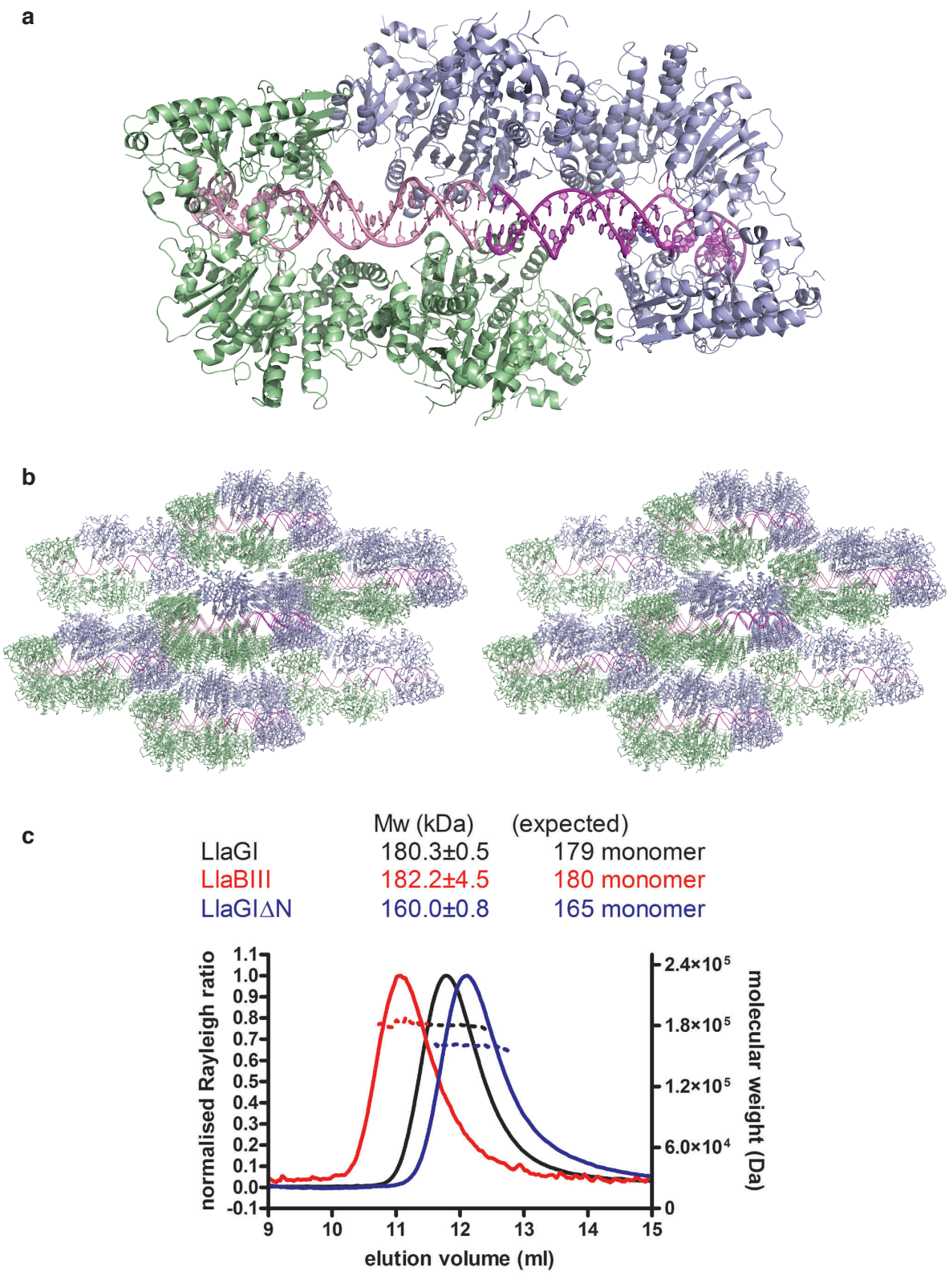
Phone: +44 117 331 2158
email: Mark.Szczelkun@bristol.ac.uk



Supplementary Fig. 1. The activities of a Type ISP enzyme. a, LlaBIII and LlaGI target sequences (methylated adenine in cyan). **b,** The schematic illustrates the previously-suggested mechanisms of action of the two disparate activities of the Type ISP enzyme, i.e. DNA modification and restriction, based in part on the mechanism proposed for the heteropentameric Type I enzymes.[18,19,20,60]

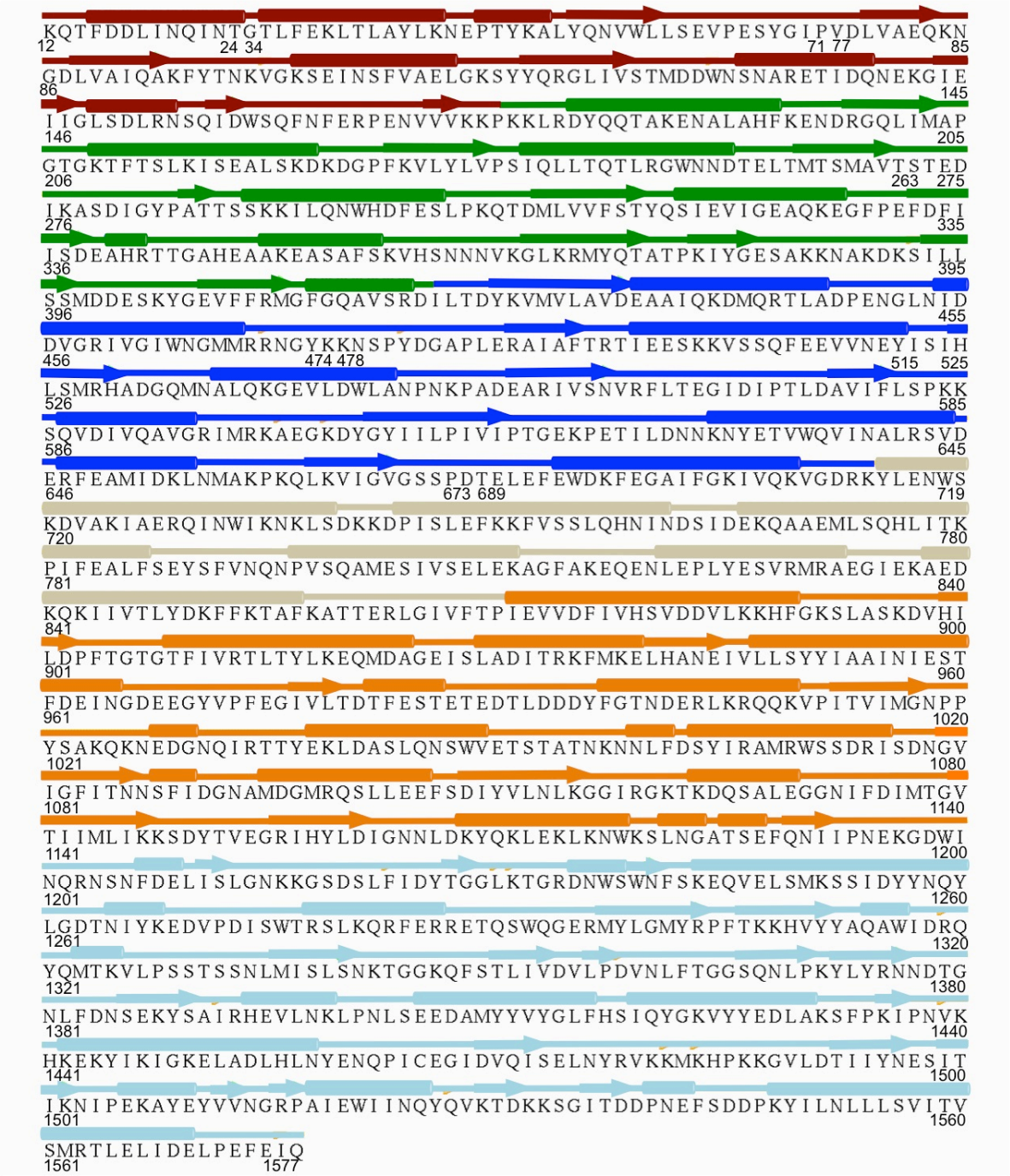


Supplementary Fig. 2. Protein-DNA interactions. A schematic diagram of the LlaBIII-DNA interactions. Amino acid colours are according to domains in Fig. 1a. DNA bases are cyan (A), brown (T),blue (C) and yellow (G), main chain interactions are pink lines, side chain interactions are purple lines, and boxes represent R1119 (black), M1137 (tan), F1134 (orange), Y1021 (lavender). The target adenine is +1, with downstream positions defined as positive and upstream positions defined as negative.

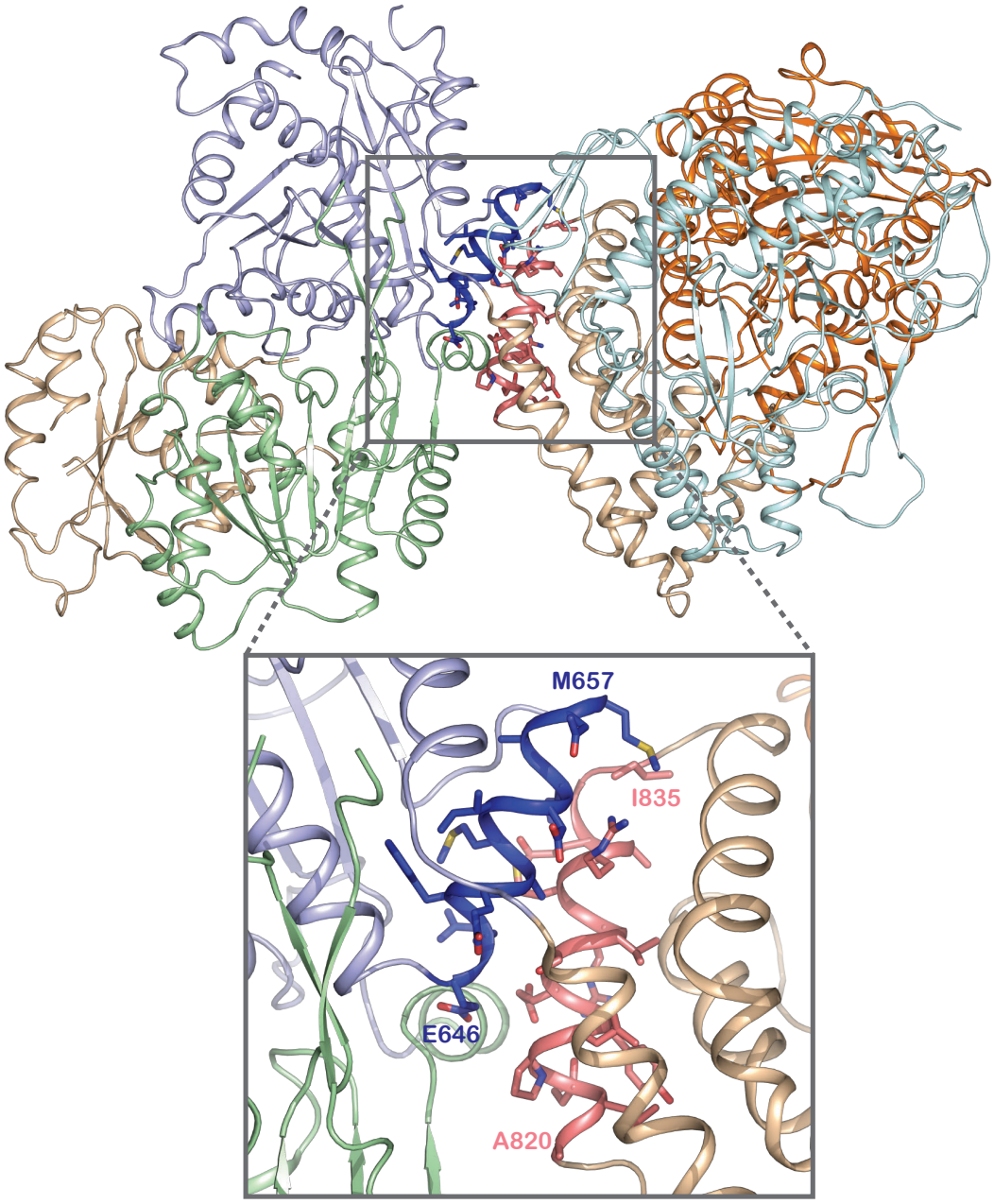


Supplementary Fig. 3. Type ISP enzymes are monomeric. a, Ribbon diagram of the two molecules of LlaBIII-DNA in the asymmetric unit of the

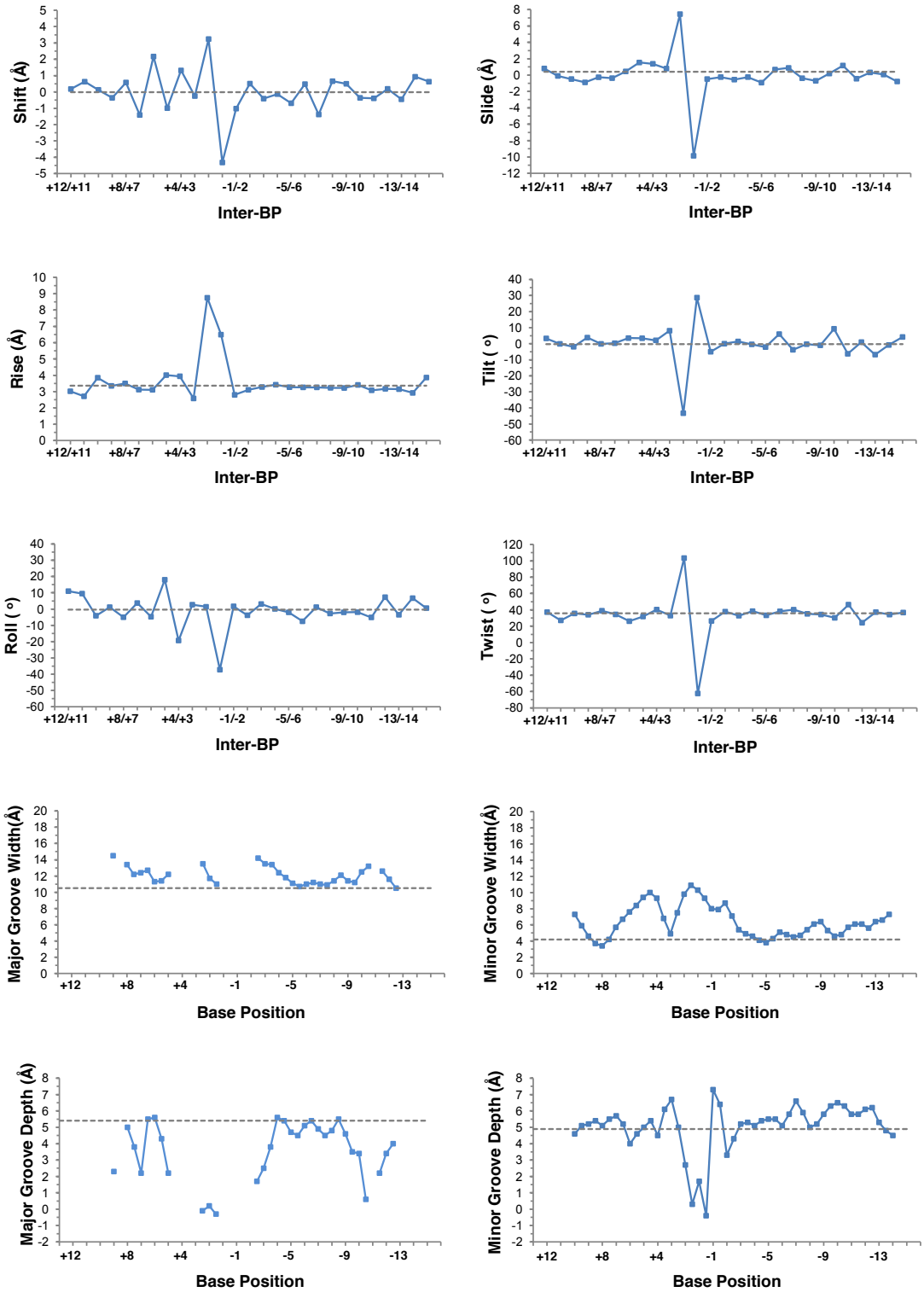
crystal. **b**, Stereoview of the crystal packing in the unit cell of the LlaBIII-DNA crystal. **c,** SEC-MALS traces of 0.2 mg/ml of apo-LlaGI, LlaBIII or LlaGIΔN. Solid lines represent light scattering (left *y*-axis) whilst dotted lines show the calculated molecular masses (right *y*-axis). We note that despite the similarities between LlaGI and LlaBIII, the retention time on the Superdex 200 column was different, suggestive of different conformations of the apo-enzymes.



Supplementary Fig. 4. Secondary structure representation of LlaBIII. The secondary structural information was obtained using the program STRIDE.[61] The cylinders represent helices and the arrows represent strands. The structural elements are coloured according to the domains as in Fig. 1a.

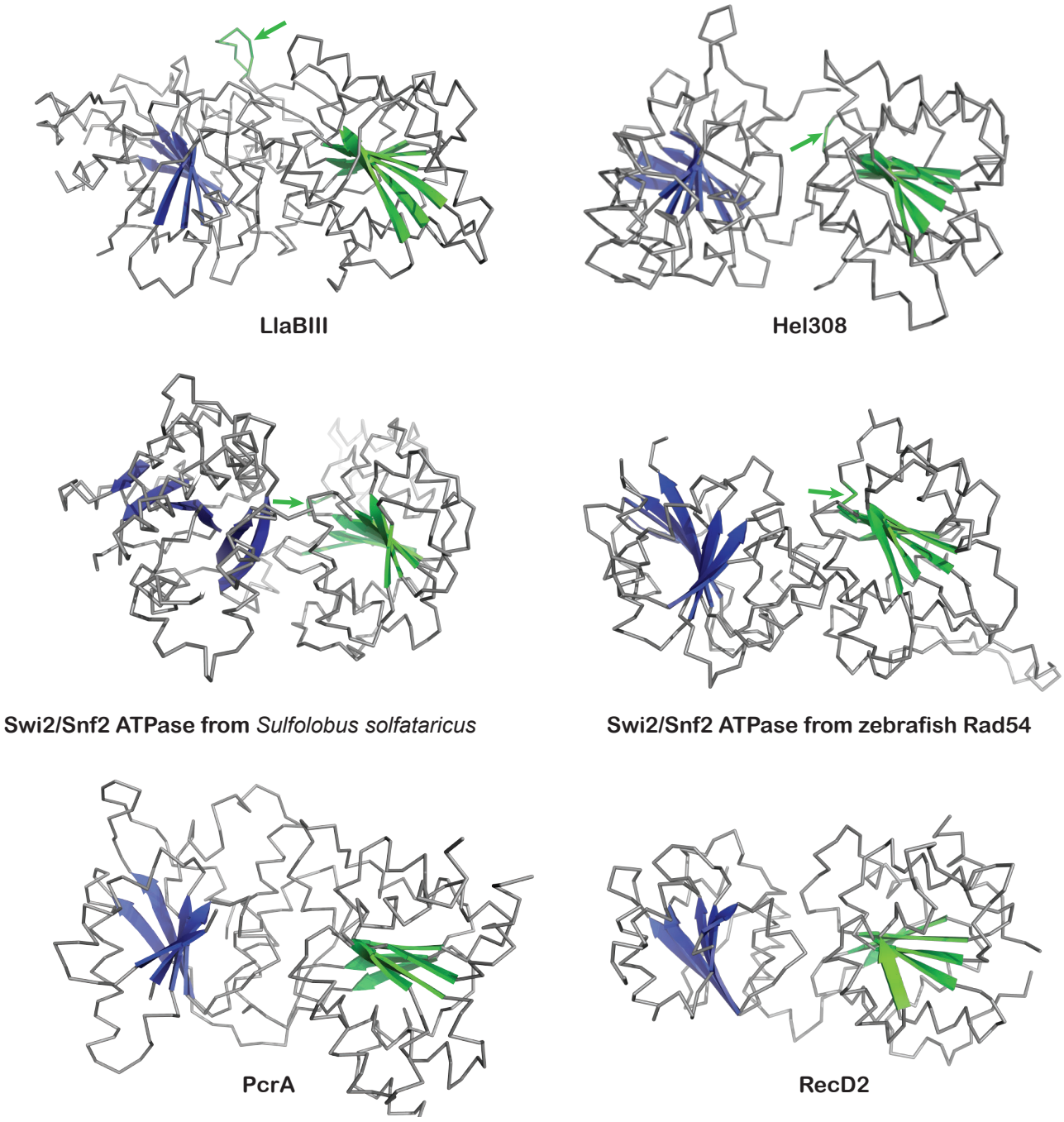


Supplementary Fig. 5. The hinge about which the nuclease-ATPase move about the coupler. A zoom in view of the two laterally packed α-helices (E649-N656 and P823-I835), one from the C-core (blue) and one from the coupler (wheat), about which the nuclease-ATPase pivot are displayed in the inset. The side chains of the residues in the helix are displayed as sticks.



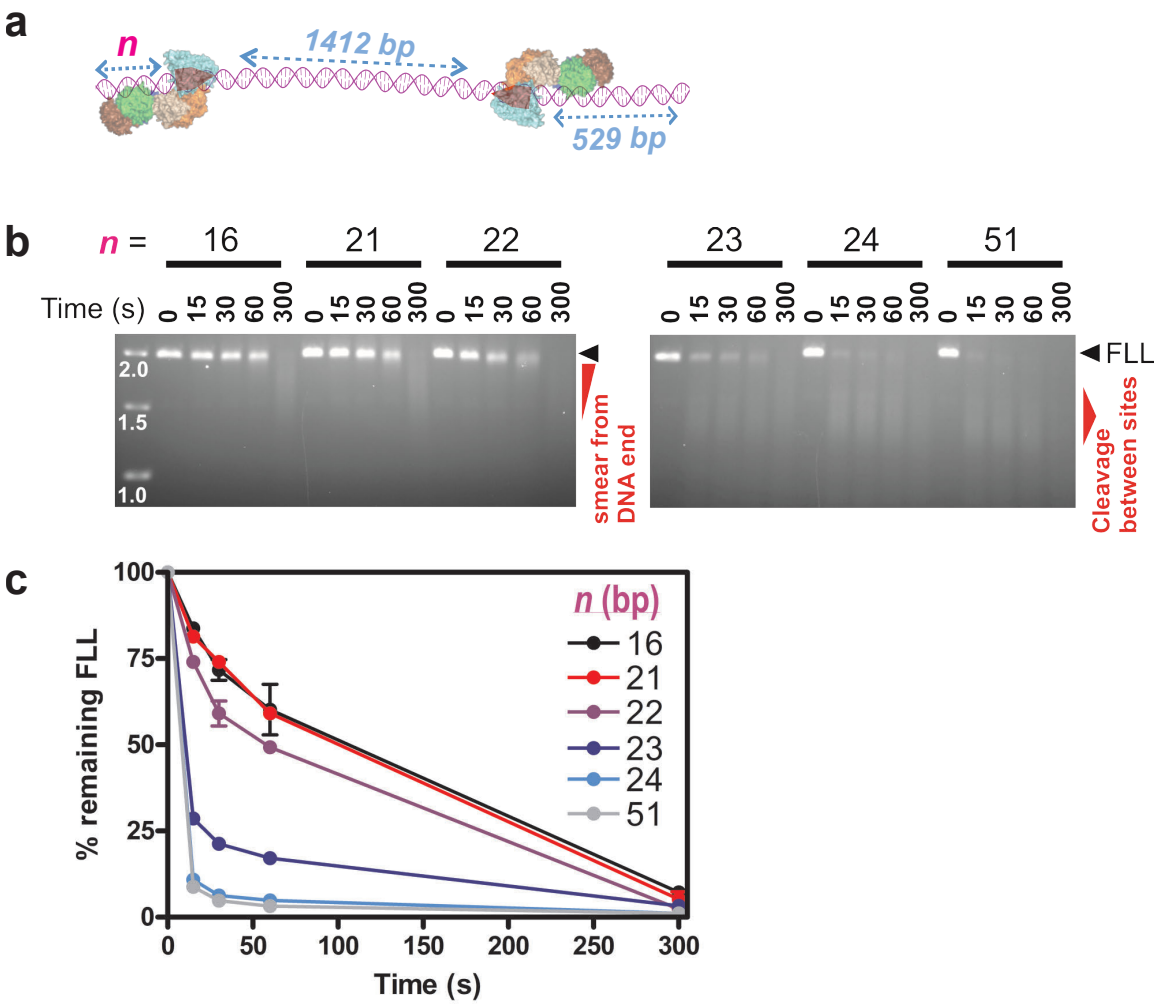
Supplementary Fig. 6. Geometrical parameters of the DNA bound to LlaBIII. The geometrical parameters were calculated for the DNA chains E and F using the program Curves+.[62] The major groove width and depth at

some of the regions could not be calculated due to large deformation. The grey lines indicate the average parameter values for a B-DNA obtained by Curves+ using the coordinates 1BNA.[62]



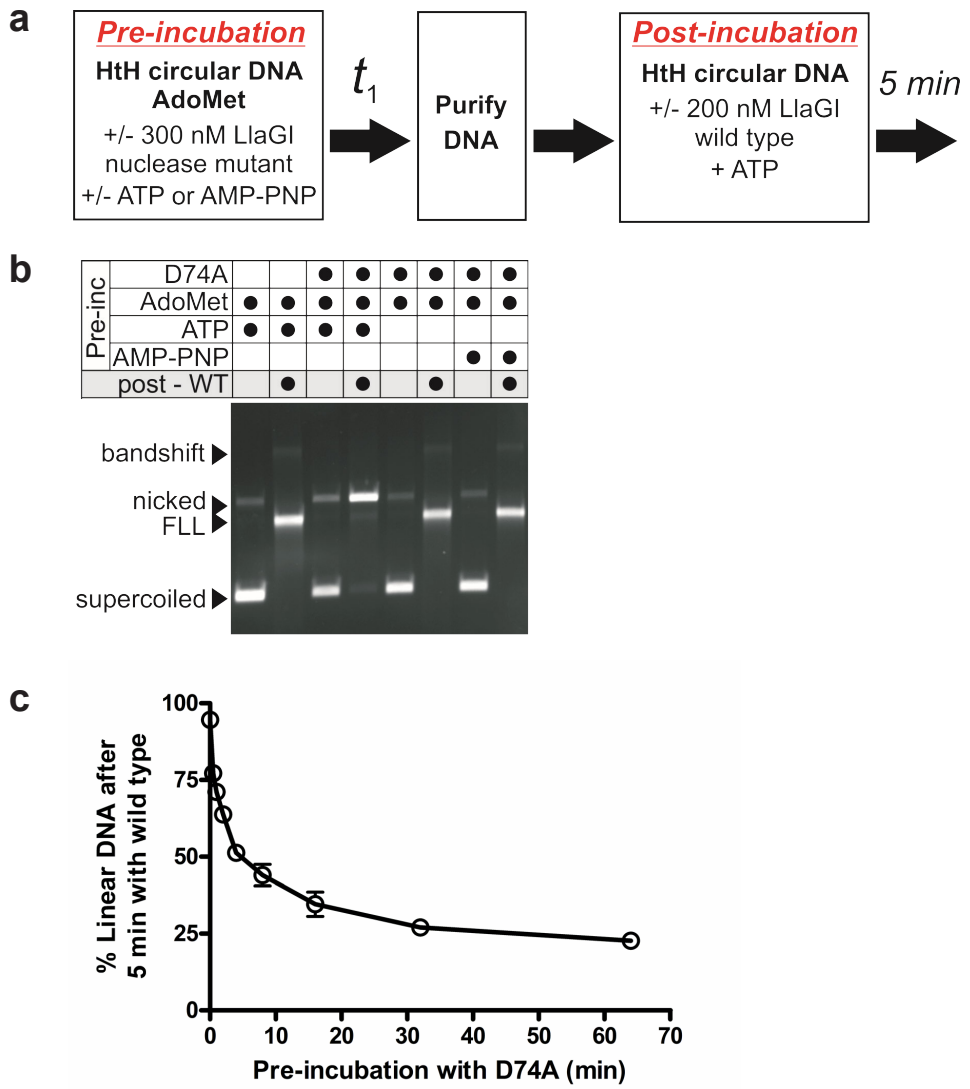
Supplementary Fig. 7. Relative orientation of the N-core and the C-core domains in SF2 and SF1 helicase-like ATPases. Ribbon diagrams highlighting the relative interdomain orientation of the N-core (green) and the C-core (blue) domains of LlaBIII, dsDNA unwinding SF2 helicase Hel308[28] (PDB code 2P6R), the SF2 helicase-like dsDNA translocase Swi2/Snf2

ATPase from *Sulfolobus solfataricus*[31] (1Z63), Swi2/Snf2 ATPase from zebrafish Rad54[14] (1Z3I), SF1 helicase PcrA[26] (2PJR) and RecD2[29] (3GP8). For clarity, only the central β-sheets of the two core domains are shown in ribbon while the other structural elements are displayed as C^{α} trace. The interdomain orientation of LlaBIII is similar to that of the DNA bound Hel308, PcrA and RecD2, and the apo structure of the Swi2/Snf2 ATPase from zebrafish Rad54, suggesting that the structure of the LlaBIII ATPase is in an active conformation. In this orientation, the helicase motifs are positioned to bind to ATP (Fig. 3a). The β-hairpin loop (S383-D390), which is unique to LlaBIII, and the equivalent regions in other SF2 ATPases are coloured green and indicated by an arrow.



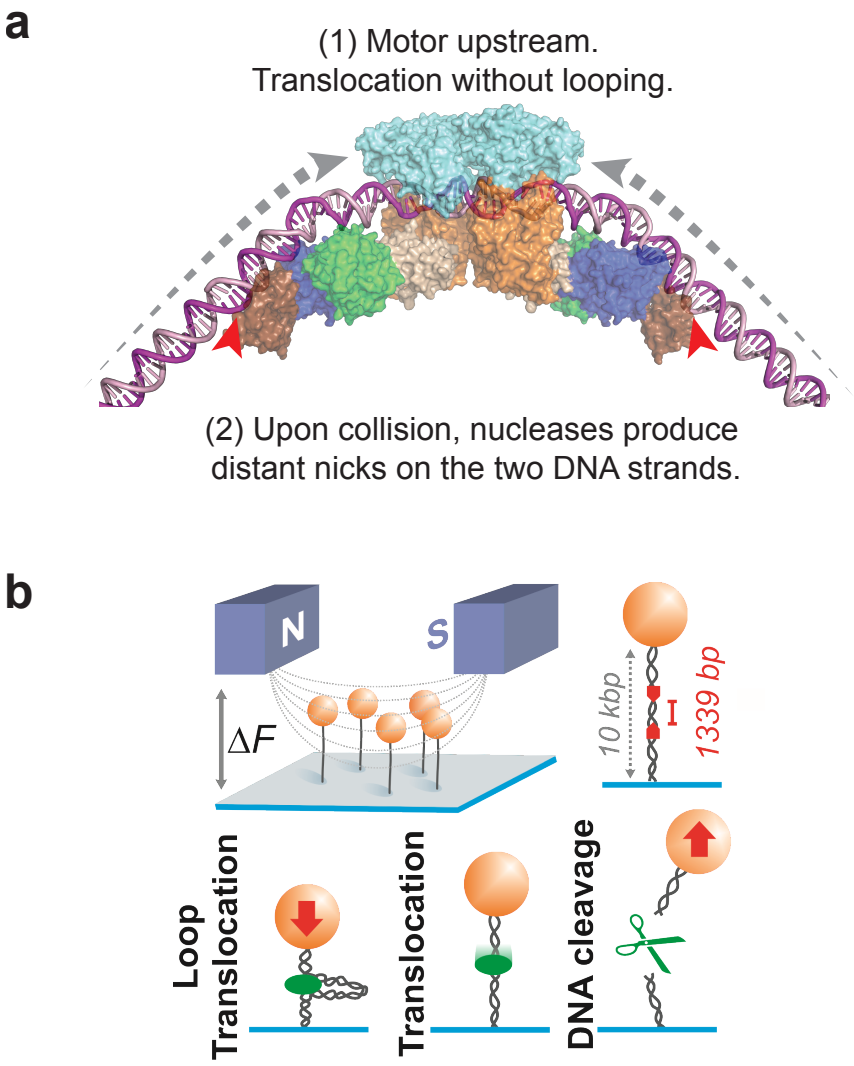
Supplementary Fig. 8. The effect of upstream DNA on DNA cleavage. a, Cartoon of the linear DNA used in the cleavage assays. The DNA upstream of the left-hand site was varied (*n*) and the effect on DNA cleavage observed. **b,** Agarose gel showing time points from cleavage reactions at the time points indicated (after addition of ATP to initiate cleavage), using Full Length Linear (FLL) DNA in which *n* varied as indicated. With 51, 24 or 23 bp upstream, both enzymes initiated translocation and DNA cleavage occurred rapidly, midway between the sites. With 22 bp or less upstream, the DNA cleavage was slower, and originated from the DNA end. This was due to the translocating enzyme from the right-hand site colliding with the left-hand

enzyme at its site, followed by a slow initiation of the left-hand motor, causing movement of the collision complex and processing of the DNA end. **c,** Graph showing the quantitation of the FLL DNA cleavage as indicated. Error bars are the standard deviation of three repeats (and are smaller than the points in some cases).

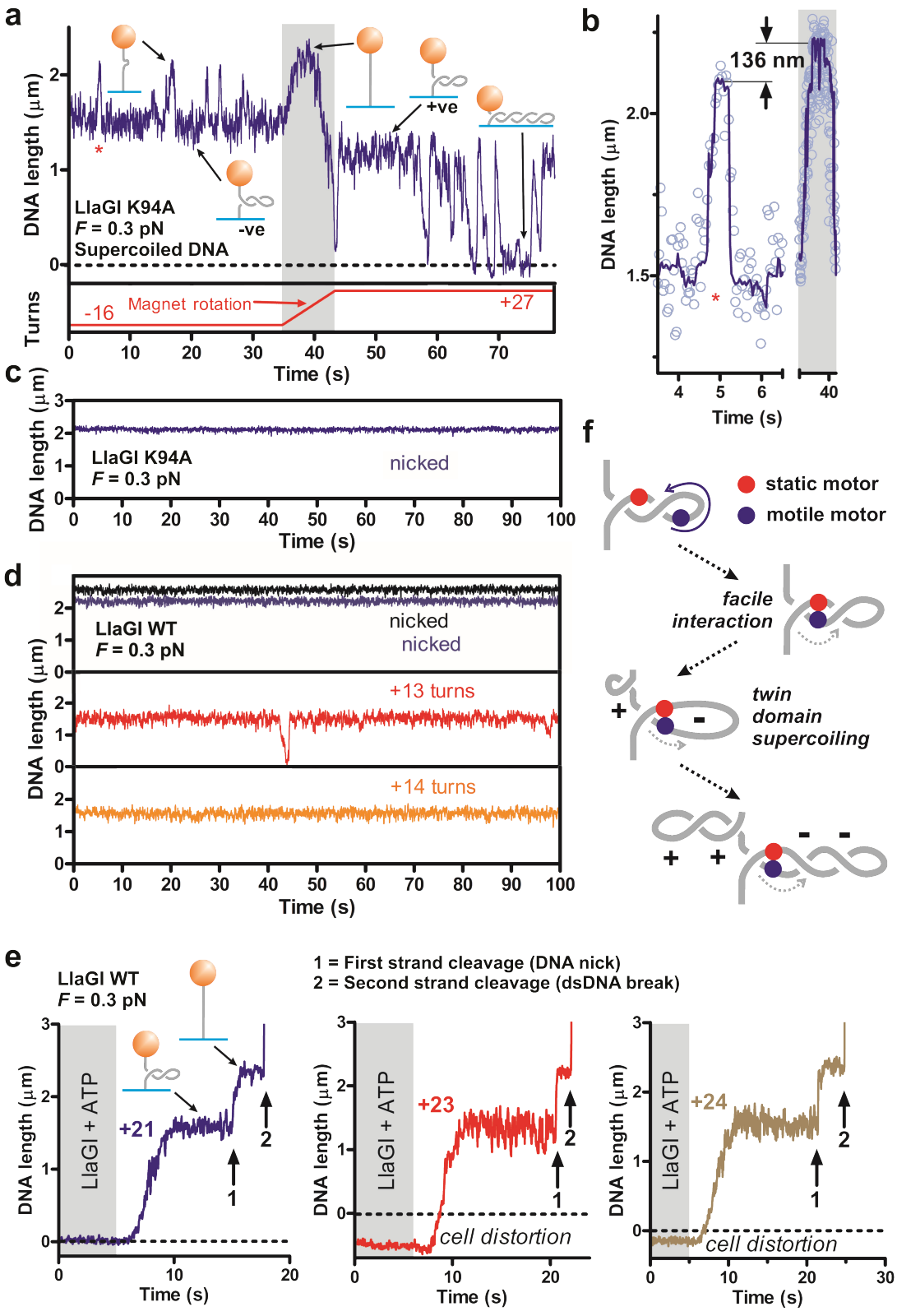


Supplementary Fig. 9. DNA cleavage inhibition assay as a measure of DNA modification. a, Reaction scheme. Methylation of at least one site on the head-to-head DNA will block formation of full length linear (FLL) DNA. **b,** Agarose gel DNA following preincubation (t_1) for 60 minutes as shown. AdoMet was added at 100 μM and was supplemented after 30 min. Nucleotides were added at 4 mM. Nicking (and trace FLL) in lane 4 results from incomplete methylation.[6] Bandshifts are due to stable LlaGI interactions with the product DNA. Linear changes to brightness/contrast were applied to clearly show the faint DNA bands. **c,** Time course of DNA modification. DNA

was pre-incubated for the times shown with LlaGI D74A, 100 μM AdoMet and 4 mM ATP. Purified DNA was treated with wild type LlaGI and 4 mM ATP for 5 minutes. Error bars are the standard deviation of three repeats (and are smaller than the points in some cases). Under the same conditions, long-range translocation, formation of the collision complex and the first DNA strand break is 100% complete within ten seconds (Supplementary Fig. 14d).

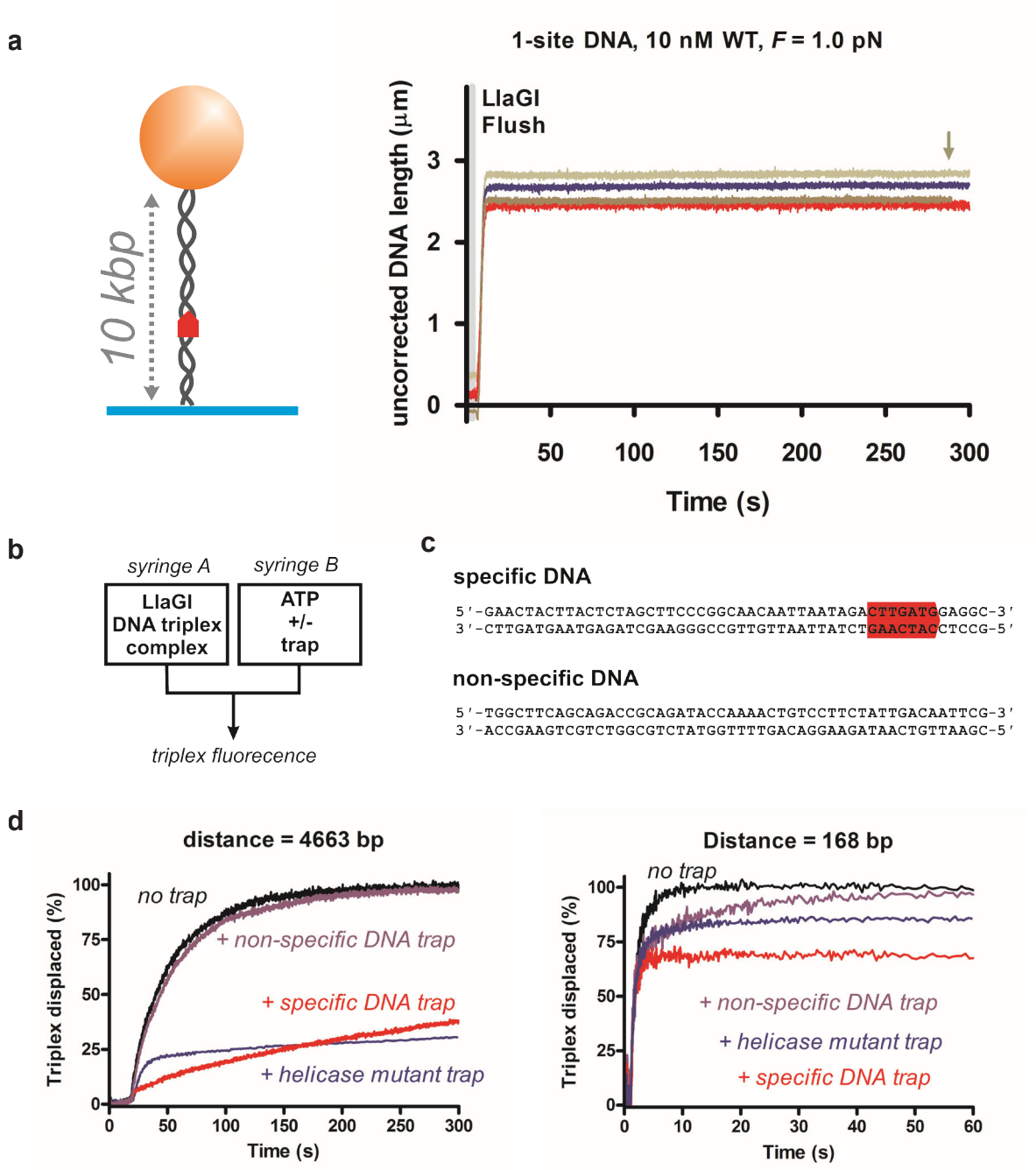


Supplementary Fig. 10. Magnetic tweezers assay. a, Head-to-head collision model following loop-independent translocation places the nucleases ~75 bp apart. **b**, Magnetic tweezers assay using multiple DNA-paramagnetic bead tethers. Topologically unconstrained head-to-head DNA were stretched at different forces using magnet above the flow cell.[32,56] Different translocation models can be distinguished by effects on DNA length. dsDNA breaks produce instantaneous loss of bead-tracking.



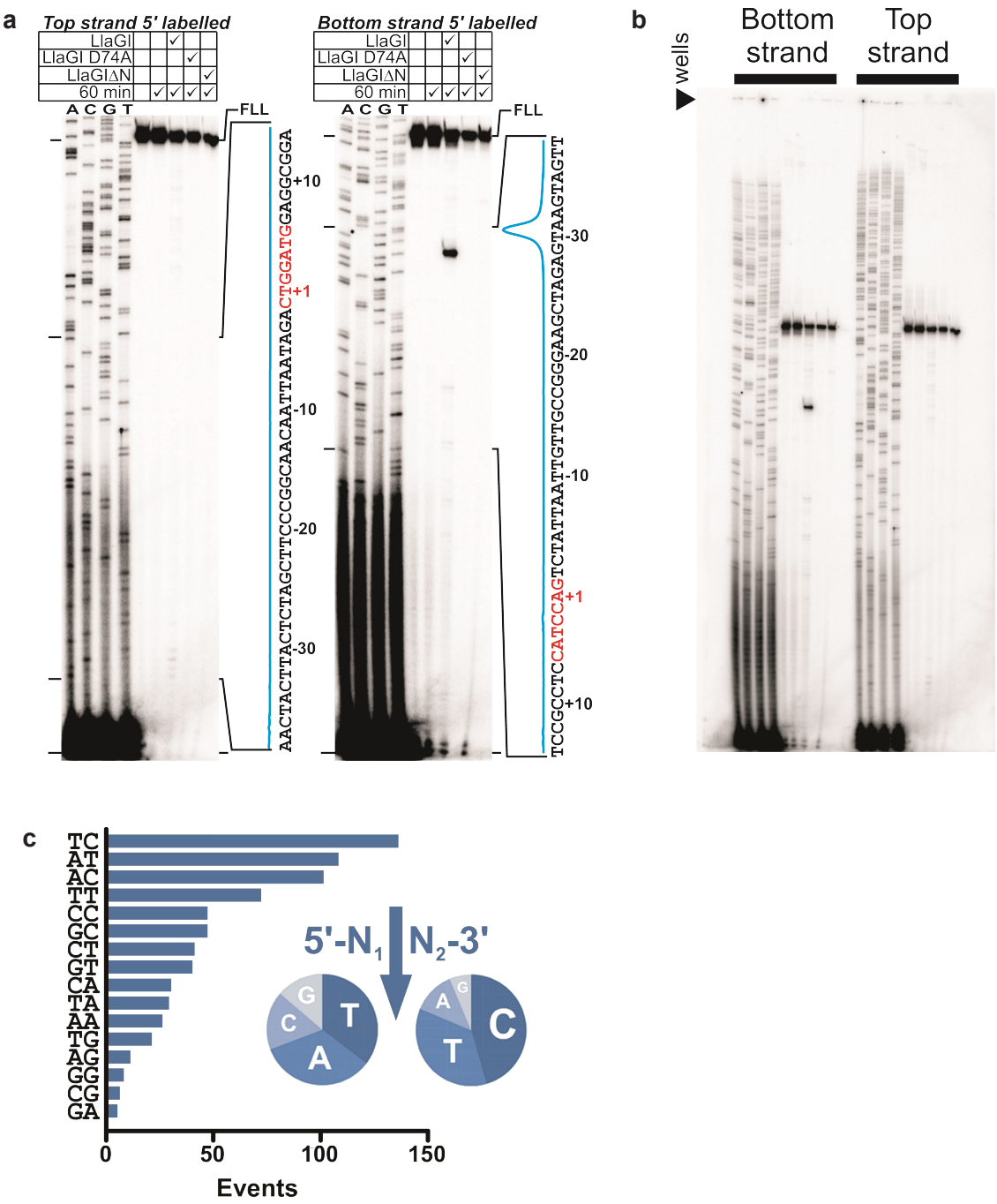
Supplementary Fig. 11. Evidence for loop translocation events on supercoiled DNA. a, Changes in DNA topology by a LlaGI nuclease mutant

on one-site topologically-constrained DNA (Supplementary Fig, 12a). The DNA was twisted using the magnets to form negative (-16) or positive (+27) supercoils. **b**, Detail of event indicated in panel a by an asterisk. **c**, Changes in bead height are not observed using the LlaGI nuclease mutant on one-site topologically-unconstrained DNA. **d**, Wild type LlaGI produces rare events on some positively supercoiled one-site DNA, but events on topologically-unconstrained DNA are not observed. Apparent DNA lengths in the middle and lower panel are for the plectonemically supercoiled DNA. DNA nicking by the wild type enzyme has not occurred and this would be observed as an increase of length (see panel e). **e**, Three examples of DNA cleavage of topologically constrained, positively supercoiled head-to-head DNA. Wild type LlaGI and ATP were added as in Fig. 4. The first DNA strand cleavage (event 1) releases the mechanically-introduced supercoils causing an increase in bead height to relaxed DNA. The second DNA strand cleavage (event 2) cuts the DNA, releasing the bead. **f**, Model for trapping of topological domains.



Supplementary Fig. 12. Evidence for loop-independent translocation and release of the target. a, Activity of LlaGI on one site DNA in the magnetic tweezers. Using a topologically-unconstrained one site DNA, LlaGI and ATP were flushed into the cell for 5 s. After the flow was stopped, the beads returned to the stretched DNA height. Four DNA-bead tethers are shown. One

DNA was cut after ~289 s. The other DNA were intact after 300 s. **b**, Stopped flow syringe arrangement for the translocation trap experiments. **c**, Sequences of oligoduplexes used as traps. **d**, The effect of traps on triplex displacement at two different spacings (distances are between the LlaGI site and the fluorescent triplex).

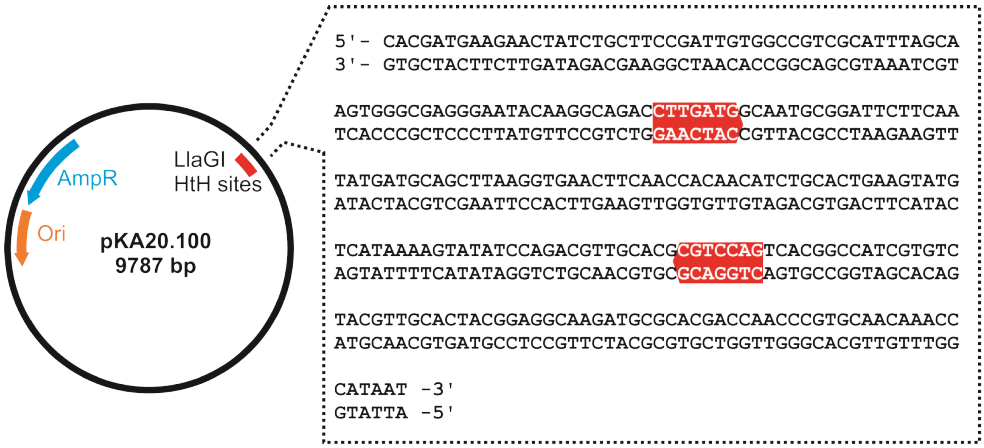


Supplementary Fig. 13. a, Denaturing PAGE mapping of ATP-independent strand-specific DNA nicking by LlaGI following one hour incubation with a DNA (Supplementary Fig. 14a) radiolabelled on either the top or bottom strand. **b,** Full scan of gel used for analysis in panel a. **c,** Dinucleotide cleavage frequency for all time points. (*inset*) independent cleavage frequencies.

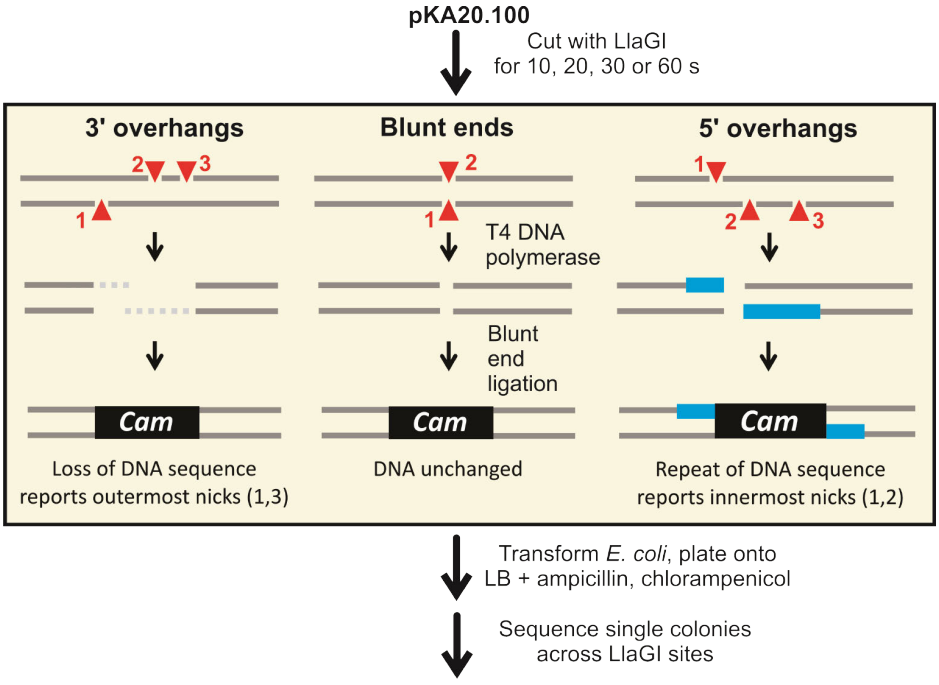
a

```
5'-AGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTAC
3'-TCGTTACCGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATG

TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGT
AGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCA

TGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGC
ACGTCCTGGTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAATAACG

TGATAAATCTGGA-3'
ACTATTTAGACCT-5'
```

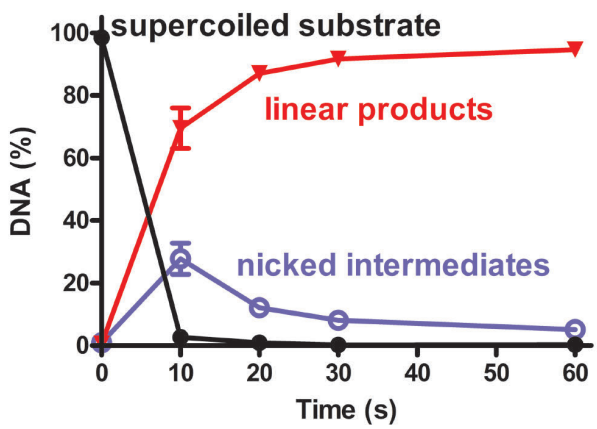
b



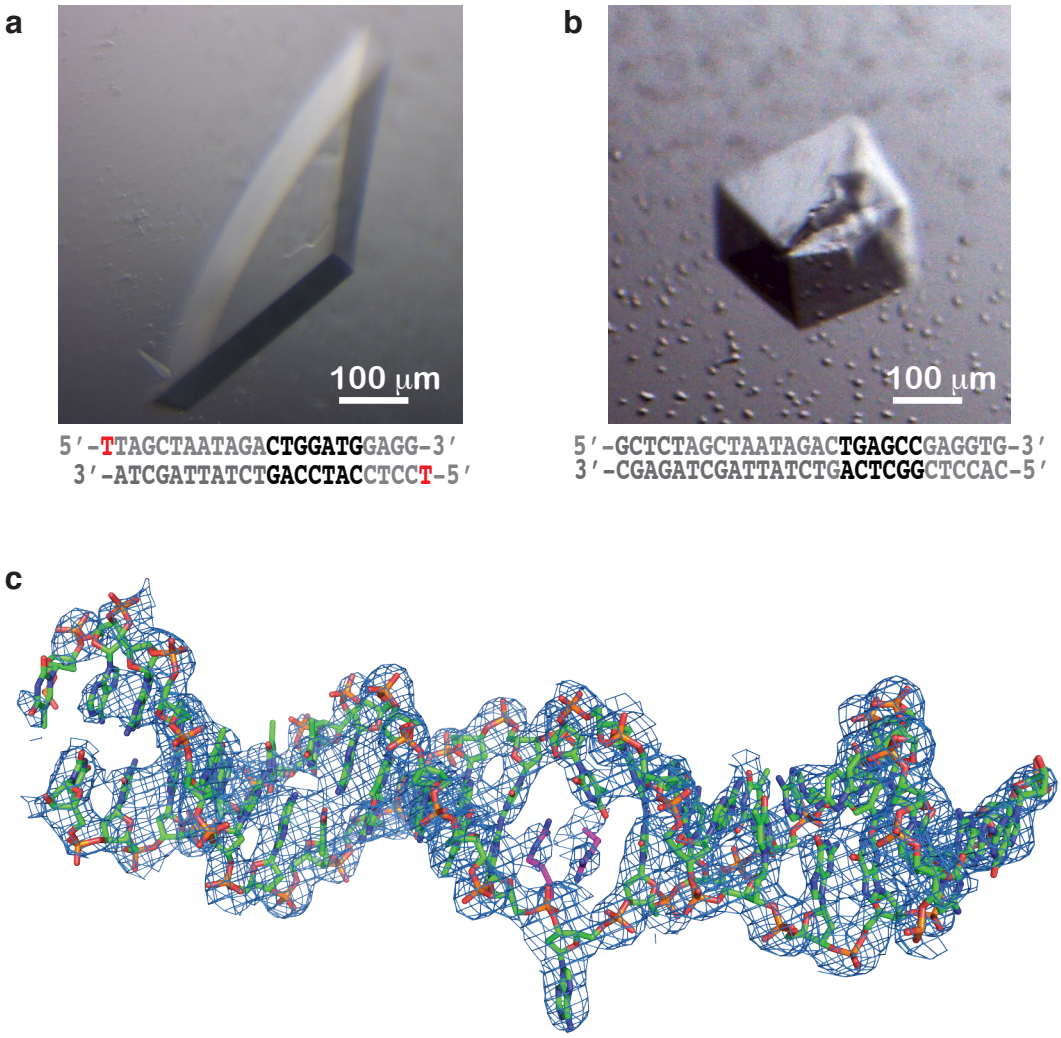
c



d



Supplementary Fig. 14. Single DNA cleavage event mapping method. a, DNA sequence used in the nicking assay in Supplementary Fig. 13a. **b,** Plasmid pKA20.100, with the local sequence around the head-to-head LlaGI sites shown. **c,** Cleavage event mapping method. Following cleavage by LlaGI at defined time points after initiation with ATP, the DNA was processed by T4 DNA polymerase to produce DNA ends. Depending on the location of the (multiple) nicks, this will result in loss of DNA sequence (for 3' overhangs), gain of DNA sequence (for 5' overhangs) or no change (for blunt ends). Because of the nature of the T4 DNA polymerase processing, the methods reports on the outermost pair of nicks for 3' overhangs and the innermost pair of nicks for 5' overhangs. A blunt-ended chloramphenicol resistance cassette was ligated into the DNA, *E. coli* cells transformed, and the cells plated onto LB plus antibiotics to select single colonies with chloramphenicol resistance. DNA was extracted and sequenced across the LlaGI sites to give the location of the chloramphenicol resistance cassette and the adjacent DNA ends. **d,** Cleavage time course for reactions used in the single cleavage event mapping (Fig. 5b). pKA20.100 and LlaGI were mixed and the reaction started with 4 mM ATP. DNA at each time point were separated by agaorse gel electrophoresis, and quantified. Error bars are the standard deviation of three repeats (and are smaller than the points in some cases).



Supplementary Fig. 15. Crystals and electron density map. a, Crystal of LlaGIΔN bound to a 22 bp DNA with a one-nucleotide 5'-overhang, marked in red. **b,** Crystal of LlaBIII bound to a 28 bp DNA. **c,** A simple Fourier electron density map of the DNA bound to LlaGIΔN^{Se} at a contour level of 1σ calculated using SAD phases.

Supplementary Results

Supplementary Table 1. Data collection and refinement statistics

	LlaBIII-DNA	LlaGIΔN^{Se}-DNA (Friedel pairs unmerged)
Data collection		
Space group	P1	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	62.9,112.7,146.2	87.8,222.6,117.6
α, β, γ (°)	103.3,90.9,105.8	90.0,105.3,90.0
Resolution (Å)	50.0-2.70 (2.77-2.70)	50.0-3.20 (3.28-3.20)
R_{sym}	5.6 (66.8)	5.5 (43.7)
I / sI	11.2 (1.2)	11.6 (1.8)
Completeness (%)	96.4 (96.3)	97.0 (97.5)
Redundancy	1.8	2.0
Refinement		
Resolution (Å)	48.2-2.7	
No. reflections (total/test)	99188/4969	
R_{work} / R_{free}	21.8/26.0	
No. atoms		
Protein	23312	
DNA	2283	
Ion	3	
Water	148	
Average B-factors		
Protein (Å^2)	75.2	
Ion (Å^2)	83.6	
Water (Å^2)	57.7	
R.m.s. deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.686	

Supplementary Table 2. Interdomain movement in LlaBIII

Domain	Rotation angles* between the respective domains of chain A and B of LlaBIII	Polar angles of rotation (omega, phi, kappa)
Coupler	**0°**	**0.0°, 0.0°, 0.0°**
Nuclease	**7.2°**	**68.6°, 24.6°, 7.2°**
N-core domain of ATPase	**4.6°**	**81.3°, 26.7°, 4.6°**
C-core domain of ATPase	**3.3°**	**82.8°, 26.2°, 3.3°**
Nuclease-ATPase	**4.5°**	**77.5°, 24.3°, 4.5°**
Methylase	**1.1°**	**85.8°, 140.4°, 1.1°**
TRD	**1.4°**	**107.8°, 131.4°, 1.4°**

*The rotation angles were obtained by superposing a domain of chain B of LlaBIII onto the corresponding domain of chain A. Prior to this, chain B was superposed onto chain A with respect to the coupler domain (residues 715 to 870) to ensure a common frame of reference (see Fig. 1b). The structures were superposed using the Secondary-structure matching method in Coot.[51]

Supplementary Table 3. DNA primers used in the study

Primer	Sequence
KA050F	GTGAACGGAAGAACACAGCGAAAC
KA082F	GCACCAATAACTGCCTTAAAAAAATTACGC
KA082R	GCGAAAATGAGACGTTGATCGGC
KA40F	CTGACAGCCCATTCATCAATTCTTTGCAG
KA41F	GTTATCGACGCGTCGTCAGCCAGGGATG
KA42F	CGTAGATAAAAAATGCTGCGTCAATAGAGAGGAAG
KA69F	CACGCGTCCATTCACGGCCATCGTGTCTAC
KA69R	GTAGACACGATGGCCGTGAATGGACGCGTG
KA43F	GCTTTGCATCGACGCAGGGTCACTTG
KA44F	CCCGGCGGCATATTGATGAGAGCTTAC
KA45F	CAAATGGATGCGGTCCAGGAGGAGAAAATG
KA46F	GTGGCACTAACATCTATGTATGAGGTAGAGGGCTG
KA47F	CCTCCATCCAATCTATTAATTGTTGCCGGGAAGCTAGAG
ms293SpeF	GCGTAAGTCTCGAGAACTAGTAGAATAGTGTATGCGGCGACC
ms293NotR	GCGTAAGTGCGGCCGCTGACCATGATTACGCCAAGC
FOli LlaGITr-529	GAAAAGTGCCACCTGACGTCTAAGAAACC
ROli LlaGI Tr-16	GCTCTAGCTAATAGACTGGATGGAGGCGCTAGAGGATC
ROli LlaGI Tr-21	GCCAAGCTCTAGCTAATAGACTGGATGGAGGCGC
ROli LlaGI Tr-22	CGCCAAGCTCTAGCTAATAGACTG
ROli LlaGI Tr-23	ACGCCAAGCTCTAGCTAATAGACTG
ROli LlaGI Tr-24	TACGCCAAGCTCTAGCTAATAGACTG
ROli LlaGI Tr-51	CACACAGGAAACAGCTATGACCATGATTAC
FOli LlaGI-529	GAAAAGTGCCACCTGACGTCTAAGAAACC
ROli LlaGI-16	GCAACAATTAATAGACTGGATGGAGGCG
ROli LlaGI-21	TCCCGGCAACAATTAATAGACTGGATG
ROli LlaGI-22	TTCCCGGCAACAATTAATAGACTGGATG
ROli LlaGI-23	CTTCCCGGCAACAATTAATAGACTGGATG
ROli LlaGI-24	GCTTCCCGGCAACAATTAATAGACTG
ROli LlaGI-51	CTATTAACTGGCGAACTACTTACTCTAGCTTCC
ROli LlaGI int 0	CATCAAGGTCTGCCTTGTATTCCCTCGCCCAC
FOli LlaGI int +100	CAAGCAATGCGGATTCTTCAATATGATGCAGCTTAAG

Supplementary References

60. Studier, F. W., and Bandyopadhyay, P. K. (1998). Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci. USA.* **85**, 4677-4681.
61. Frishman, D. & Argos, P. Knowledge-Based Protein Secondary Structure Assignment. (1995). *Proteins: Structure, Function, and Genetics* **23**, 566-579.
62. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*. **37**, 5917-5929.